Intelligent common spoken Chinese phonetic input method and dictation machine

Publication number: CN1127898
Publication date: 1996-07-31
Inventor: LINSHAN LI (CN)
Applicant: LI LINSHAN (CN)
Classification:
- international: ***G06F3/16;*** **G06F3/16;** (IPC1-7): G06F3/16; G10L7/08
- European:
Application number: CN19951000623 19950126
Priority number(s): CN19951000623 19950126

Also published as:
CN1153127C (C)

Abstract of **CN1127898**

The input method for common speech of the Chinese language to transform pronunciation into correspondent character includes sound processing and language decoding procedures. In sound processing procedure, "hidden Markov model" and "tone model" are used and in language decoding procedure "Chinese language model" is used. A dictation machine using "intelligent learning technique" based on said method can transform input speech into text and display it.

[19]中华人民共和国专利局

[11] 公开号 CN 1127898A

[12] 发明专利申请公开说明书

[21]申请号 95100623.1

[51]Int.Cl[6]

[43]公开日 1996 年 7 月 31 日

G06F 3/16

[22]申请日 95.1.26

[71]申请人 李琳山

地址 中国台湾

[72]发明人 李琳山

[74]专利代理机构 中国专利代理(香港)有限公司

代理人 曾济洪 萧掬昌

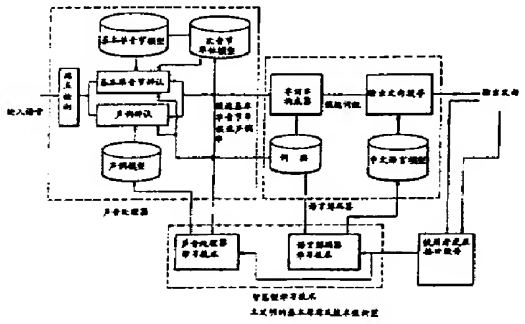
G10L 7/08

权利要求书 8 页 说明书 24 页 附图页数 12 页

[54]发明名称 智慧型国语语音输入方法及国语听写机

[57]摘要

一种国语语音输入方法，用以将任意文句的国语语音直接转换成相对应的中文文字。该方法包括声音处理过程以及语言解码过程两大部分，声音处理过程中利用了“隐藏式马可夫模型”和“声调模型”；语言解码过程中以“中文语言模型”，找出所对应的中文字。一种国语听写机，用以根据上述方法而将输入的语音转换成文字显示出来，此国语听写机尚包含许多“智慧型学习技术”，使得这套听写机更具备不时“学习”的“智慧”。



权　　利　　要　　求　　书

1. 一种国语语音输入方法，用以将任意文句的国语语音直接转换成相对应的中文文字，该方法包括声音处理过程以及语言解码过程两大部分，其特征在于，该声音处理过程是利用针对国语音节特性所发展的以“次音节单位”为基础而产生的“隐藏式马可夫模型”加以组合的“基本单音节模型”以及针对声调特性的“声调模型”来对比输入国语语音的每一音节以及声调的机率，进而辨认之；该语言解码过程针对该声音处理过程送来的一连串音节，以“中文语言模型”，找出所对应的中文字。

2. 一种国语语音输入方法，用以将任意文句的国语语音直接转换成相对应的中文文字，该方法包括声音处理过程以及语言解码过程两大部分，其特征在于，该声音处理过程是利用针对国语语音特性所发展的“次音节单位模型”及针对声调特性的“声调模型”直接与输入的语音对比，再由对比所得的“次音节单位模型串”及“声调串”中找出对应的音节，进而辨认之；该语言解码过程针对该声音处理过程送来一连串音节，以“中文语言模型”，找出所对应的中文字。

3. 根据权利要求1 或2 所述的方法，其中该“中文语言模型”是以“字”、“词”或“词群”为单位做统计分析或参酌词类、语法、语意分析获得的信息或规则等语言学知识，然后将结果适度组合。

4. 根据权利要求3 所述的方法，其中该“词群”是将某些特性相类似的词组合而成，这些特性包含同一个字结尾，同一个字起头的特性、文法特性、语意特性以及统计特性。

5. 根据权利要求1 或2 所述的方法，其中该声音处理过程包含下列步骤:

(1) 对输入的国语语音做端点检测，找出声音由那一点开始，那一点结束；

(2) 对输入语音做基本单音节及声调辨认，该基本单音节的辨认是以基本单音节模型或“次音节单位模型”与输入语音作对比找出对应的基本单音节，该声调辨认则自一声调模型中找出对应的声调，进而串接成词或句；和

(3) 以前述基本单音节及声调辨认所获得的可能基本单音节及声调中可能性及出现机率较高的基本音节串和声调串，作为候选基本音节串和候选声调串输出至语言解码器。

6. 根据权利要求1 所述的方法，其中该“基本单音节模型”是以“次音节单位模型”为基础而建立，并进而串接成词或句。

7. 根据权利要求2 或6 所述的方法，其中是以“受后接韵母起始音素影响的声母”和“不受前后音影响的韵母”为该“次音节单位”。

8. 根据权利要求2 或6 所述的方法，其中该“次音节单位”为“受后接音素影响的音素”。

9. 根据权利要求1 或2 所述的方法，尚包含一声调辨认步骤，该声调辨认是采用适用于国语连续音中的声调变化而建立的“受前后音影响的声调模型”，此模型判断每一声调受前接声调及后接声调的影响改变特性的情形，并合并接近的状况，而使所有的175 种声调模型数目大幅减少，即可完全辨识。

10. 根据权利要求5 所述的方法，其中该“次音节单位模型”和“声调模型”都是以“内插训练法”训练而成的“隐藏式马可夫模型”，其中该“内插训练法”是指在训练的第二阶段的递回训练过程中，每一次递回训练得到的模型被与第一阶段的模型进行某种程度的“内插”，以充分利用第一阶段模型的精确度，加速第二阶段的训练，使得所需要的训练语料可以适度减少。

11. 根据权利要求6 所述的方法，其中该“次音节单位模型”是以“内插训练法”训练而成的“隐藏式马可夫模型”，其中该“内插训练法”是指在训练的第二阶段的递回训练过程中，每一次递回训练得到的模型被与第一阶段的模型进行某种程度的“内插”，以充分利用第一阶段模型的精确度，加速第二阶段的训练，使得所需要的训练语料可以适度减少。

12. 根据权利要求5 所述的方法，其中该“基本单音节辨认”及“声调辨认”包含“连续音节比对法”和“词汇音节比对法”。

13. 根据权利要求12所述的方法，其中该“连续音节比对法”包含下列步骤:

(1) 利用输入语音音段的瞬间能量及音节长度上下限找出每一个可能的音节起始点及终点;

(2) 以“动态规划法”(Dynamic Programming) 就每一个可能的音节起始点及终点利用“次音节单位模型”或“基本单音节模型”及“声调模型”，找出整个音段中最可能的“基本单音节串”及“声调串”的组合;

(3) 以步骤(2) 的方法自整个音段的起始点开始逐步计算所有可能的单音节的起点、终点，并累积它们的分数; 和

(4) 将分数较高的音节串输出。

14. 根据权利要求12所述的方法，其中该“词汇音节比对法”是将电脑内建的词典中所有的词根据其基本单音节(没有区别声调) 或单音节(有区别声调) 的顺序建立一个“树状词典结构”; 视该树状结构中的每一节点都是一个基本单音节或单音节，沿该树状结构往下趟到底就可以得到一个词，而后“词汇音节比对法”是利用此一词典结构中，每一基本单音节或单音节在每一词中与前后基本单音节或单音节相连的情形，优先考虑最可能相连的基本单音节或单音节; 因而

大幅减少搜寻空间并提高正确率。

15. 根据权利要求14所述的方法，尚可依词出现的频率找词，即越常用到的词越优先考虑。

16. 根据权利要求1 或2 所述的方法，其中该语言解码过程包含下列步骤，

(1) 根据该声音处理过程送来的一系列候选基本单音节串及候选声调串，在一“字词串构成器”中与电脑内建立的词典对比，将可能的同音字或所对应的各个可能的同音词找出来，产生候选词组；和

(2) 以“中文语言模型”计算该候选词组中各个词连成句出现的统计机率并参酌语言学知识以最可能的句子作为输出答案。

17. 根据权利要求16所述的方法，其中该“中文语言模型”所计算的各个词连接成句的统计机率，是由单独一个“字”、“词”或“词群”出现的机率及两两相连或三个相连，或共同出现在同一句中的统计机率来计算。

18. 根据权利要求1、2、4或17 所述的方法，其中该“词群”的分类方式包含下列步骤:

(1) 以语言学分析的词类、语法、语音知识将所有的词分成词类一致，语法、语音一致的若干群;

(2) 把步骤(1) 分好的词类，语法、语意一致的每一个词群中的词，根据大量文字资料库中的统计特性(亦即前接什么词，后接什么词，和什么词共同出现在句子中等) 进一步分成统计特性一致的小词群；和

(3) 再利用统计特性，将步骤(1)，(2)中因词类、语法、语意不同而分开的小词群，但事实上统计特性很接近的，再把它们合并起来。

19. 根据权利要求15所述的方法，其中该“中文语言模型”中的语言学知识是以语言学对中文词类、语法、语意分析所获得的知识、

规则或信息，并得与前述统计所获得的语言信息相结合的方式构成。

20. 根据权利要求16所述的方法，其中自该“声音处理过程”送来的每一候选单音节都存有在该声音处理过程中所辨认的分数，分数高的单音节所构成的字或词应予以优先考虑。

21. 根据权利要求16所述的方法，其中该“中文语言模型”亦计算每一单音节出现的频率、两个单音节两两相连出现的频率、三个单音节相连出现的频率等，依此类推。

22. 根据权利要求16所述的方法，其中该“中文语言模型”尚可用于部分更正声音处理部分的错误。

23. 一种训练国语语音辨认系统迅速学习新使用者的声音的训练方法，用以训练一国语语音听写机迅速学习辨认新使用者输入的国语语音，其特征在于，该方法包含下列步骤:

(1) 以很多位不同的语音发出的声音来训练国语每一个“次音节单位”以及“声调模制”的“隐藏式马可夫模型”，因许多位说话者声音特性各不相同，故常需很多个高斯机率混合才能描述每一个状态;

(2) 以某一个新使用者所发出的训练语音中的“次音节单位”音段，自上述很多位使用者的“隐藏式马可夫模型”的许多高斯机率混合中找出最接近该某一新使用者声音的那几个高斯机率混合，而把其他的高斯机率混合抛弃，即建立出此一新使用者的“隐藏式马可夫模型”;

(3) 当新使用者继续发出同一“次音节单位”音段时，此一新发的“次音节单位”音段的特性就可以再平均进入在步骤(2) 所求得的新使用者的“隐藏式马可夫模型”中，算出新的高斯机率混合，而得到新的“隐藏式马可夫模型”; 和

(4) 重覆步骤(3) 的方法，新使用者的声音在“隐藏式马可夫模型”中的成分于是越来越多，即可得更精密的描述新使用者声音的“

隐藏式马可夫模型”。

24. 根据权利要求23所述的方法，该方法尚包含随时于电脑银幕上线上更正电脑辨别错误的音的步骤，并将此结果立即送入一存储器中，并当场重复上述的步骤(2)、(3)和(4)，使得机器亦即时学到新的声音，下次再辨认就用新的模型，故正确率可以不断提高。

25. 一种国语听写机，用以听写国语文句，其特征在于，包括一滤波及类比／数位转换器，以将语音输入信号滤波及转换为数位信号，一个人电脑及附加数位信号处理电路板用以接收该转换送来的数位信号而加以处理，一特征求取器及一基频检测器与该个人电脑相连接用以检测及计算由该个人电脑所收到的数位信号的基频及其他多种特征，一隐藏式马可夫模型处理器，配合高斯机率混合处理器，以计算每一段语音音段的端点，并辨认其基本单音节及声调，一以“字”、“词”或“词群”为基础来计算统计机率并参酌语言学知识的中文语言模型处理器，以计算输入语音音节的各个同音字、词的机率，并进而组成字词串或字句，并将辨识结果送回该个人电脑，一训练学习装置用以训练和学习出所有“次音节单位”、基本单音节及声调的“隐藏式马可夫模型”的机率数值以及“中文语言模型”的机率数值或知识，然后将此数值或知识送入该个人电脑。

26. 根据权利要求25所述的国语听写机，其中语音输入是以语音音段(单音节、词、音韵段或整句话)为单位。

27. 根据权利要求25所述的国语听写机，另外包括一荧光幕用以显示输入的注音符号及中文文字以及方便的改正错误的软件，使得使用者可以直接用滑鼠在荧光幕上改正错误，完全不需用到键盘。

28. 根据权利要求25所述的国语听写机，其中尚包含一动态存储装置，用以暂存使用者的语词和习惯用语或所输入的某一段文字中反覆出现的特别语词，并根据该语词的出现频率，储存于不同的存储器

中，这些语词及其信息可以并入听写机的整体词典及中文语言模型中，也可以在事后清洗掉。

29. 根据权利要求25所述的国语听写机，其中尚包含一常用词存储器和一罕用词存储器，该听写机操作时原则上只在该常用词存储器内找词，找不到时才到该罕用词存储器内寻找，并将找到的罕用词存入该常用词存储器内；该常用词存储器内储存的常用词若久不使用，即移入该罕用词存储器中。

30. 一种训练国语听写机学习适应新使用者声音及环境的方法，其特征在于，该方法包含数种学习方式，其中：

(1) 第一种学习方式是以分阶段的“学习例句”来阶段性的自动学习使用者的声音；

(2) 第二种学习方式是使机器自动“线上”即时学习使用者的声音，此方式可配合第一种学习方式随时线上学习使用者的语音；

(3) 第三种学习方式是线上自动学习环境噪声；和

(4) 第四种学习方式是线上自动学习使用者的用字、用词及构句习惯。

31. 根据权利要求30所述的方法，其中该分阶段的“学习例句”包含数段学习步骤，每一段步骤须由新使用者念一组经特别设计的例句，该组例句不但以最少的字句包含所有国语语音的基本单位音(例如次音节单位、音素、声母、韵母、单音节等)，并使常出现的单位音多出现几次，故多念几次，可以把“隐藏式马可夫模型”训练得更精确，藉着反覆练习该组例句，而使该国语听写机习惯新使用者各种发音方式，并将该发音方式记录起来，而各阶段“学习例句”中基本单位音出现的重点不同，故可以分阶段以最快的速度提高辨认新使用者声音的正确率；例如第一阶段是以最少的字句把最重要的基本单位音念好，使机器以最快的速度初步学会新使用者的声音，而其后各阶

段再逐步尽快提高辨认正确率。

32. 根据权利要求30所述的方法，其中该线上学习步骤可在做学习训练时或正式使用国语听写机期间进行，使用者随时更正该国语听写机所显示辨认错误的声音或文字，使该听写机随时学习正确的语音及语词，并将更正的语音对应文字内容储存起来。

33. 根据权利要求30所述的方法，其中该自动学习环境噪声的步骤是与权利要求23所述的(3)、(4)两步骤所描述的学习新使用者的声音的步骤同时进行，让新使用者的环境噪声也自动被平均进去成为“隐藏式马可夫模型”的成分，以使该国语听写机熟悉学习环境的噪声。

34. 根据权利要求30所述的方法，其中国语听写机学习新使用者声音的“学习例句”，是由电脑由语料库中选出，是先将所有的国语基本单位音给予不同的分数，同一句子中所包含的不同基本单位音愈多，则其分数愈高，愈会优先选出，并利用一参数描述各个基本单位音出现的频率分布，故可使用此参数做为选句的基础。

35. 根据权利要求30所述的方法，其中该第四种学习方式是一方面动态调整“中文语言模型”中的统计数值及语言学知识并可在词典中加入新词，一方面将使用者的语词和习惯用语或所输入的某一段文字中反覆出现的特别语词暂存于一优先被选取的动态存储装置中，并根据该词语的出现频率，储存于不同的存储器中。

说　　明　　书

智慧型国语语音输入方法及国语听写机

本发明涉及一种智慧型国语语音输入方法及国语听写机。本发明为同一发明人的台湾专利申请案第82106686号的改良，利用经改良方法得使利用国语语音输入中文文字的方法更为方便好用且更为精确。

目前中文电脑的输入方法百家争鸣，或用注音，或用字根，或用笔划，但没有一种是众所公认最好的，因为没有一种真正最方便。这是因为有的输入速度较慢，有的需要特别训练，有的方法特别要背口诀，久了不用会忘掉等，而从从都会、不需训练的注音符号法，则因其速度太慢，而无法通行。在众多中文输入法中，速度最快的是仓颉法、大易法或类似的方法，但此方法却只有专业人员在长期训练下才会用，一般人不常用就会忘掉。事实上，这是现阶段我国社会资讯化最大的障碍，因为“中文输入”变成一种专门职业，一般人自然不会常用它。这些方法不方便的基本原因，是尝试把中国字转成几个按键，由键盘输入；但事实上键盘是西方拼音文字下的产物，中国文字不是拼音文字，所以由键盘输入就自然不方便了。

既然键盘输入不方便，还有什么其他方法可用呢？很多人很早就想到了可用声音输入。只是用声音输入的技术困难太多，几乎是不太可能的事，所以一直没有这方面的方便产品问世。技术上困难的原因有三：(1) 需要辨认的字汇太大了，中文常用字至少五千个，常用词至少十万个，这种数字已超出技术可行的范围；(2) 中文字的同音字太多，即使知道是什么音，又如何能方便而快速无误的知道是什么字呢？(3) 要能“即时”听写国语，就必须在极短时间内解决如此困难

的问题，更是不容易。

发明人发明的第82106686号专利申请案基本上已可以解决上述困难，是因为：(1) 选用国语单音节为电脑处理的基本单位：中文字、词的数目虽大，不同的单音节却只有约1300个，是语音辨认技术上可以克服的范围；知道是什么单音节以后，可以再由其前后的单音节去判断可能构成什么词、什么句。(2) 藉助“中文语言模型”，可以靠大量的训练文字资料，统计出每一个字或词的前后与其他不同的字或词衔接的机率，由这些机率可以算出当一个音节前后与其他音节衔接时，这些音节最可能是代表什么字，这种方法可以大部分解决同音字的问题，不能解决的再生荧光幕上予以更正。

本发明中，就是在前项发明的架构下，再进一步发展出两项更完善的新技术：(1) 以“次音节单位”(次音节单位，sub-syllabic units，指比音节更小的声音单位，声母、韵母、或“音素(phoneme，如子音、母音等”为基础，经特殊训练(如“内插训练法”) 所产生的“隐藏式马可夫模型(Hidden Markov Models)”，以及考虑连续国语语音中声调特性变化的“声调模型”，并辅以“连续音节比对法”及“词汇音节比对法”，来进行更完善的国语单音节的辨认；如此单音节的辨认技术将不仅可以有效辨认“断开的单音节”，更可以相当精确的辨认“连续音中的单音节”，故使用者的输入语音将不再限制是一连串的“断开的单字(单音节)”，也可以是“断开的词(多字词时各字音间是连续不断开的)”、“断开的音韵段(音韵段，prosodic segment，为一个或若干个词构成的，是人在说话时一口气告一段落时自动断开的音段，音段内各字音是连续不断开的)”、甚或是“整句完全连续的”国语语音。(2) 以大量中文文字资料中统计出字与字、词与词前后相连或同时出现的机率信息，辅以中文语言学对中文词类、语法分析所获得的知识或规则所建构成更完善的“中文

语言模型”，加上更有效率的搜寻法，可以在所辨认出来的可能的国语单音节中，更迅速而正确地找出所代表的同音字。这两项技术都是针对中文及国语的特性发展出来，结合起来以后，可以精确的辨认“连续音中的单音节”，使使用者输入的语音型态可以更为方便自然而且多元化；而同时所需的运算量并不会增加多少，而正确率却可维持同样高或更为提高。所有技术可以用软件完成，并轻易写入任何装有“数位信号处理晶片”(DSP Chip)的“数位信号处理电路板(DSP Board)”(这类晶片及电路板市面上产品很多，故很容易在不同的电路板或晶片上发展出不同的产品)，只要晶片的运算速度够快，电路板上的存储容量够大，它就能“即时”输入。这片电路板可以插入任何一台AT级以上的个人电脑上，故使用方便，价格亦可大为降低。以上述的基本技术及功能为基础，本发明又进一步发展出许多“智慧型学习技术”，使得这套听写机更具备不时“学习”的“智慧”。这包括：自动学习新使用者的声音，使得新使用者可以很快开始使用、自动学习使用者的环境噪声并适应该噪声、不断线上学习使用者的声音、用字、用词(包括专有名词)、构句等，使得正确率可以继续上升等等。所有这些都将在以下详细说明。

本发明涉及国语语音输入方法及国语听写机，该国语听写机指利用语音处理技术的方法及根据此方法研制而成的机器，可以“听写”任意文句的国语，亦即使用者对着机器说任意文句的国语，机器可以将之辨认出来，把语句转换成文字，显示在荧光幕上(以中文文字)。其主要应用是作为中文电脑的输入。就好比有一个“听写员”，听了使用者的语句，并将之输入电脑。当然，在输入电脑之后，就可以加以任何处理、修改、编排、储存、印出、传递到远方等应用。简言之，这种机器使中文电脑“会听国语”。这种“国语听写机”和一般看到的能辨认国语语音的机器系统最大的不同有二：(1)它必须能“听写”

由极大字汇(中文常用词至少10万以上，常用字至少5 千以上) 组成的任意文句，因为一般电脑要输入的中文可以是任意的文字。(2) 它必须快到可以“即时”(Real-time) 辨认，完成听写，亦即使用者不能在说完话后慢慢等中文字显示，因为一般电脑输入的应用都是即时的，这两个不同点使得“国语听写机”在技术上易做到，故到目前为止尚没有真正可以有效使用的产品出现。目前各研究单位所发展的“国语语音辨认系统”，或者只能辨认少数的有限字汇(例如100 个地名等) ，或者正确率仍很低尚不便于使用等，均与本发明不同。

因为上述“国语听写机”在技术上十分困难，本申请案的发明人早在1989年就提出第一项申请案，当时的发明是将上述构想再增加一些条件，使上述构想在以下三个条件下，在技术上变成可行，可以确实作到: (1) 机器只适应会听特定语音的声音: 亦即一架机器一次只听一个使用者的声音，每个使用者在购买机器时可以对机器说一番话作成“训练资料”，输入机器后机器可以调适到听懂他的话，换使用者时只要换一套“训练资料”即可，并不构成太大困难，因为这种机器一次只有一个人在用。发音不正确的人也可以用不正确的发音去训练机器，机器基本上也可以一样听不正确的发音。(2) 输入以单音节为电脑处理的单位: 国语有“一字一音的特性”，亦即每一个字构成一个单音节，故可以先辨认出所有的单音节，再由这些单音节找出相对应的字、词及句子。(3) 输入的文字可以允许有少量的错误: 事实上任何输入法均可能输入错误的字，只要输入的文字可以先显示在荧光幕上，使用者看到有错时，可以用简单的方法，借助方便的软件予以更正。在这样的条件下，使用前项申请案中的发明，每分钟约可输入150 字，其中约有17字需要更正; 由于更正的软件十分方便，每分钟的“净输入”可达约110 字。若使用本发明，则效果会更好。需要说明的是，目前中文输入法中最快的方法也可达到约每分钟110 字以

上，不过全台湾只有少数专业人员在长期练习下才能达到。使用本发明则任何人均可随时达到这个数字。

因此本系列的研究发明，自1989年的第一项申请案开始，就是使任何会说国语的人，在不需训练及永不忘掉的情况下，方便又快速使用本发明所述的中文语音输入方法及根据此方法所制成的国语听写机来输入中文。

本发明的其它目的和优点可由下列较佳实施例配合附图的说明叙述如下，其中：

图1 为本发明的基本原理与技术架构。

图2 为两种可能的“次音节单位”举例，一种以“声母、韵母”为基础，一种以“音素”为基础，并以“电脑”一词的基本单音节“ㄉㄧㄢ，ㄋㄠˇ”为例说明。

图3 为考虑在连续国语语音中声调特性受前后音影响有所改变的“声调模型”举例说明。

图4 为“连续音节比对法”的说明图例。

图5 为“词汇音节比对法”中所用的“树状词典资料结构”。

图6 为发明人于1989年所申请的第78105818号案中的“以字为基础的马可夫中文语言模型”。

图7 为发明人于1993年所申请的第82106686号案中的“以词为基础但以字来计算的马可夫中文语言模型”。

图8 为结合统计特性及词类语法语意等语言学知识或规则来作“词群”分群方法的举例说明。

图9 为在本发明的技术下各种可能的国语语音输入方式。

图10说明“语言解码器”的智慧型学习技术可能作法的细节举例。

图11说明用电脑自动选取“学习例句”的方法。

图12为本发明的一个较佳具体实施例。

本发明的基本原理及架构，请见图1，分为“声音处理器”以及“语言解码器”两个部分，另外包括“智慧型学习技术”。第一部分针对输入的语音信号，以声音处理的方式负责辨认出是那一连串的单音节；第二部分则针对辨认出来的一系列可能的候选单音节，以语言解码的方式负责找出各是那一个字。在第一部分“声音处理器”中，则先对每一输入语音音段（可以是单音节、词、音韵段或整句话）检测出其端点，再分别进行“基本单音节辨认”（“基本单音节”是指不考虑声调的，例如辨认出为“ㄉㄧㄢ”）及“声调辨认”（例如辨认出其为“第四声”），则可知其为那一个（或一串）音节（例如“ㄉㄧㄢˋㄋㄠˇ”等）。这些辨认出来的音节串就都被送到“语言解码器”之中去找出正确的同音字，首先先由“字词串构成器”由词典中把所有可能的同音字或同音词都找出来。再藉助有效的搜寻法，使用一套完善的“中文语言模型”找出机率最大的（或最可能的）同音字或词串作为输出。

如果输出不正确，使用者可以在荧光幕上予以更正。更正后不仅输出的文句可以改正，改正的信息也同时进入“智慧型学习技术”的部分；其中“声音处理器学习技术”可以进一步改正“基本单音节辨认”所用的“次音节单位模型”及“声调辨认”所用的“声调模型”，而“语言解码器学习技术”可以进一步改正“词典”及“中文语言模型”，使整个系统更能适应使用者的声音及用词、构句等。

首先说明本发明在图1中第一部分“声音处理器”的第一步工作，也就是端点检测法。这是作语音辨认的人所熟知的技术。基本上所有声音一输入，先由取样器对其波型取样，变成一串数据，即可输入电脑。电脑即可根据这些数据计算其“瞬间能量”（即短瞬间能量有多大）及“过零率”（即单位时间内波形由正变到负通过“零”的次数），根据这两种数据，电脑即可判断声音由那里开始到那里结束，

其余是噪声，可以去除。例如韵母的能量比噪声高很多，声母有时能量不高，但过零率比噪声高很多，故根据这两者即可把噪声和声音分开来，再就声音部分加以辨认。

其次说明“声音处理器”中的“基本单音节辨认”部分，国语单音节共约1300个，如果扣除四声变化，则只有约四百多个“基本单音节”（例如“ㄅㄚ”、“ㄅㄚˊ”、“ㄅㄚˇ”、“ㄅㄚˋ”、“ㄅㄚ·”当成5个单音节，则共有约1300个，当成1个“基本单音节”，则共有约四百多个），本发明将四声分出来单独考虑，故先当成共有四百多个基本单音节来辨认；经多年来深入研究，发现以本发明所发展出来针对国语音节特性的以“次音节单位”为基础，经特殊训练产生的“隐藏式马可夫模型”，或者再进一步加以组合成为“基本单音节模型”来作对比，效果最佳。这是因为国语单音节中混淆音组极多（例如“ㄅㄚ”、“ㄆㄚ”、“ㄇㄚ”、“ㄈㄚ”、“ㄉㄚ”、“ㄊㄚ”、“ㄋㄚ”、“ㄌㄚ”、“ㄍㄚ”、“ㄎㄚ”、“ㄏㄚ”、……都非常接近），正确无误的辨认将十分困难；上述特殊方法为本发明在台大发展出来，针对国语音节特性所找出的方法。

图2举两种可能的例子说明以“次音节单位”为基础来建立“基本单音节模型”并进一步串接成词或句的情形。图2(a)中使用“受后接韵母起始音素影响的声母”和“不受前后音影响的韵母”为“次音节单位”。传统上，国语的400多个基本单音节可以分解成声母／韵母(INITIAL／FINAL)的格式，例如“ㄉㄧㄢ”“ㄋㄠ”中，“ㄉ”、“ㄋ”为声母，“ㄧㄢ”、“ㄠ”为韵母，其中共有约22个声母和41个韵母。一般而言，声母比较短、能量比较小、较不稳定，因此很容易受到后接韵母的影响，相对的，由于韵母一般较

长、能量较高，因此较不易受到前接声母的影响。又由于国语的音节特性明显，因此可以假设声母并不太受前一个字的韵母的影响，而韵母也不太受到后一个字的声母的影响。所以，在这一个例子中，所采取的“次音节单位”是“声母”和“韵母”，但“声母”要考虑后接的韵母，亦即同一声母若后面接不同的韵母就算不同，例如“ㄉㄧㄢ”和“ㄉㄨㄢ”算用了两种不同的声母“ㄉ(ㄧ)”和“ㄉ(ㄨ)”，分别是接“ㄧ”，和接“ㄨ”的“ㄉ”。“韵母”则完全不考虑前后接的音一样不一样。此外，由于韵母通常是由数个“音素(phoneme)”组成，例如“ㄧㄢ”是由“ㄧ”、“ㄝ”、“ㄋ”三个音素构成，根据我们的观察，声母受到后接韵母的影响主要来自后接韵母的第一个“起始音素”，例如“ㄉㄧㄢ”和“ㄉㄧㄥ”的声母几乎可以是相同的“ㄉ(ㄧ)”，虽然它们的后接的韵母“ㄧㄢ”和“ㄧㄥ”不一样，但这两个韵母的“起始音素”是相同的“ㄧ”。在这样的构想设计下，就可以把国语的声音中共选出113 个“受后接韵母的起始音素影响的声母”，以及41个“不受前后接音影响的韵母”，加起来共154 个“次音节单位”；这些“次音节单位”就可用以组成共400多个国语基本单音节。这样我们一面尽可能考虑到前后音对中间音的特性的影响，一面又尽可能使模型的总数不会太多，对后面说到的模型训练有所帮助。例如图2 中“电脑”一词的两个基本单音节“ㄉㄧㄢ”和“ㄋㄠ”，在(a) 中“ㄉㄧㄢ”就由“ㄉ(ㄧ)”和“ㄧㄢ”二个次音节单位组成，“ㄋㄠ”就由“ㄋ(ㄚ)”和“ㄠ”两次音节单位组成等等。其中后者由于“ㄠ”的组成音素是“ㄚ”、“ㄨ”二者，故其声母是“ㄋ(ㄚ)”和“ㄋㄢ”的声母一样，因为“ㄢ”的组成音素是“ㄚ”、“ㄋ”二者。事实上，一段连续的国语语音可以看成

是一串这种“受前后音影响”的“次音节单位”所组成的，所以可以用这些“次音节单位”拼成的“基本单音节模型”来作连续语音中的基本单音节辨认，也可以不拼成“基本单音节模型”而直接用这些“次音节单位”来和声音对比，再在对比得到的“次音节单位串”中找出所对应的基本单音节。当然这些“次音节单位”的模型也是要用使用者的声音来训练出来的；亦即使用者必须先念若干“训练语句”，语句中包含了这些“次音节单位”，再用使用者的声音中的这些“次音节单位”训练出这个使用者的这些“次音节单位模型”。

在图2(b)中则用了另一种“次音节单位模型”，是“受后接音素影响的音素”，例如“ㄉㄧㄢ”、“ㄋㄠ”两个基本单音节中，“ㄉㄧㄢ”可以分成“ㄉ”、“ㄧ”、“ㄝ”、“ㄋ”四个音素，每一音素构成一个单位，“ㄋㄠ”可以分成“ㄋ”、“ㄚ”、“ㄨ”三个音素等等；如此国语中共可找出约33个音素；但也和上述前一个例子一样，每一个音素都会受到前后音素的影响，在这个例子中是假设每一个音素只受后接音素影响，而假设前接音素的影响小到可以不计(这也是为了模型的总数不会太多，对后面会说到的模型训练有帮助)，故“ㄉㄧㄢ”是由“ㄉ(ㄧ)”(后接“ㄧ”的“ㄉ”)、“ㄧ(ㄝ)”(后接“ㄝ”的“ㄧ”)、“ㄝ(ㄋ)”(后接“ㄋ”的“ㄝ”)、“ㄋ(#)”(音节结尾的“ㄋ”，“#”表示结尾)四个次音节单位构成；而当后接音素不同就构成不同的音素时，上述国语中共约33个音素就变成约149个“受后接音素影响的音素”。我们的实验显示，这也是一套相当有用的“次音节单位”，可以用来拼成所有的400多个国语基本音节，也相当适合用来作为连续国语语音中的基本单音节辨认。事实上国语语音中可以选用的“次音节单位”显然不只是这两种；这里举的只是两个例子而已。只要作够仔细的选择并适度考虑各单位音的特性受前后接的音的影响，都可以发展出有用的

“次音节单位”来。由于这些“次音节单位”或其组成的“基本单音节模型”可以有效在连续语音中辨认出基本单音节，故输入语音也自然可以是单字、各单字连续的多字词或“音韵段”，或是整句连续的句子。

图3 简要说明为适用于连续音中的声调辨认而建立的“受前后音影响的声调模型”。虽然国语的声调只有5 种(包括一、二、三、四声及轻声) ，但是在连续语音中，声调的变化是非常复杂的，因为每一声调的特性都会因为前后接的不同的声调而不同，因此必须选取一组适当的受前后音影响的声调模型以便描述声调的复杂变化。如果考虑所有可能的声调连接情形，则需要175 种模型，包括5^3(在句子中间，五种声调前后各可以接五种声调，故有5 ×5 ×5 种) ＋5^2(在句末，五种声调在句尾，前面各可以接五种声调，故有5 ×5种) ＋4 ×5 (在句首，因轻声字不会出现在句首，故句首字只会有四种声调) ＋5 (单独念的字) 共175 种。实际上，如果仔细考虑声调的特性，这个数目是可以大大地降低的。

以图3 为例，(a) 中是一声音前接三声后接二声，表为(3)-1-(2)，(b) 是一声音前接三声后接三声，表为(3)-1-(3) ；由图中的音高曲线可以看出来，前接三声对一声的特性影响很大，但后接二声或三声如图中(a)、(b)，对一声的影响 并无太大区别，故就中间的“一声”的模型而言，这两种(a)、(b)的前后接声调的影响可以共同使用同一个模型来描述。如果把所有这些情形都考虑进去，我们的研究显示175 种模型可以减到大约23种就足以相当理想的描述声调的复杂变化了。这样模型的总数减少了，对将来模型的训练有所帮助，以下马上就会说到。

现在说明上述的“次音节单位模型”或“声调模型”的“内插训练法”。基本上这些模型都是使用“隐藏式马可夫模型”(Hidden

Markov Models)，其基本的训练方法是此一领域的工程师所熟知的。如果要用来辨认连续语音，可以先用单音节的训练语料(也就是使用者事先念好用来训练机器的声音)来做第一阶段的模型训练，产生出可以用来做单音节辨认的模型。再以这些可以作单音节辨认的模型当起始模型，用使用者念的连续语句当作训练语料来做第二阶段的训练，经过一再反覆递回的演算。就可以产生用来做连续语音辨认的模型了。不过这样的训练法在第二阶段的训练中通常需要相当大量的连续音训练语料，使新的使用者训练机器时不堪其烦，克服这个问题的方法一方面是减少模型的数目使得每个模型可以有较多的训练语料，这也是为什么在前述的“次音节单位模型”及“声调模型”中我们尽量减少每一个模型受前后音影响的变化而使模型的总数减少的原因。另一方面则是这里所说的“内插训练法”，亦即在第二阶段的递回训练过程中，每一次递回训练得到的模型就可以和第一阶段的模型进行某种程度的“内插”(“内插”相当于一种“平均”的过程，也是工程师所熟知的技术)，这样可以充分利用第一阶段模型的精确度，加速第二阶段的训练，使得所需要的训练语料可以适度减少。

图4 则说明了如何在连续的输入语音中运用上述的“次音节单位模型”或其组合成的“基本单音节模型”及“声调模型”来作“基本单音节辨认”及“声调辨认”。图4 中所画的是瞬间能量在时间轴上的曲线，其中能量较低的点(如x、y、z)就是可能的音节起点；而如果x 是一个音节起点的话，可以根据统计出来的一个音节可能长度的上限D_{max}及下限D_{min}，找出相对于这个起点x 的音节的可能终点，也就是y-1和z-1。这时就可以使用一般工程师的熟知的“动态规划法”(Dynamic Programming)，找出整个音段中最可能的“基本单音节串”及“声调串”的组合。例如假设(x、y-1)之间的一小段语音恰是一个音节，就可以拿这一小段语音和各个“次音节单位模型”或其组成的

“基本单音节模型”及各个“声调模型”对比，每对比一次可以算出一个分数，分数最高的“次音节单位串成的基本单音节”或“基本单音节模型”和“声调模型”的组合就是该一小段语音(x、y-1)最可能的单音节了。于是我们可以由整个音段的起始点开始，一路计算下来所有可能的单音能的起点、终点并累积它们的分数；例如把累计到x-1的分数，T[x-1]，加上下一小段语音(x、y-1)的最高分数，Max S(x、y-1)，就是累计到y-1 的分数，T[y-1]。如此用电脑把所有可能的音节起点、终点分别把分数从头累计到最后，就可以把分数最高的音节串找出来了，也就是辨认的答案。这就是本发明所指的“连续音节比对法”。

除了上述的“连续音节比对法”。此外，另一重要的方法是“词汇音节比对法”，也就是充分利用词典的知识来减少音节辨认时的搜寻对比对象，并提高正确率。首先先把词典中所有的词，根据其基本单音节(也就是没有区别声调) 或单音节(也就是要区别声调) 的顺序建立一个“树状词典资料结构”如图5 所示。图中是没有区别声调的情形，当然也可以是区别声调的作法。在这个树状结构中每一节点(小圆圈) 都是一个基本单音节，而沿着树枝往下走到底就可以得到一个词，例如“医生”或“台北”等。因此当前一个音节很可能是“ㄧ”或“ㄊㄞ”时，下一个音节最可能的也许就会是“ㄕㄥ”或“ㄅㄟ”等等，至少很多原来必须考虑的基本单音节都不太可能出现了；因此搜寻对比的对象就自动减少，而正确率也可以提高。反过来顺序也是一样，如果后面一个基本单音节是“ㄕㄥ”或“ㄅㄟ”，则前一个音节是“ㄧ”或“ㄊㄞ”的可能性就提高了等等。这就是充分利用词典的知识来帮助单音节辨认的“词汇音节比对法”。这里还可以把“词频”的知识也用进来，也就是越常用到的词越应优先考虑，这也可以加快辨认的速度并提高正确率。

其次说明图1 的原理中的第二部分“语言解码器”的原理。当“声音处理器”送来一系列辨认出来的候选基本单音节串及候选声调串后，“字词串构成器”首先将每一个单音节的可能的同音字或所对应的各个可能的同音词都找出来，这是靠对比词典中的字、词及“树状词典资料结构”而查出来的。需要说明的是，必然有些单音节十分混淆，不能确定，例如图6 中的“ ㄊㄧㄥ ”和“ ㄊㄧㄢ ”很像，“声音处理器”如果没有把握它一定是那一个，可以把两个同样当作候选单音节一起送过来，“字词串构成器”会把可能的“ ㄊㄧㄥ ”的同音字及可能构成的词和可能的“ ㄊㄧㄢ ”的同音字及可能构成的词都一起列出来，这时候“字词串构成器”输出的将是一个相当庞大的“候选词组”，故需要用一个相当有效的“中文语言模型”去计算。

关于“中文语言模型”，本系列研究最早的第一项发明(图6) 的建构方式如下。例如把20,000,000字的报纸新闻资料(电脑档案) 输入电脑，电脑的程式会去计算里面的字单独及相连出现的次数，例如“中”字共出现150 个，但“中央”出现32个，“中国”出现28个……等，电脑有程式根据一定的公式，即可算出各个字出现及组合的机率。当“声音处理器”送来一串音节(注音符号) 时，这个语言模型中的程式就会有一定的公式去计算每一组可能的同音字会组合成一组句子的机率。例如在图6 中“ ㄉㄧㄢˋ ”、“ ㄋㄠˇ ”、各有很多同音字，但“电脑”两字相连的可能性最大，而“ ㄍㄨㄛˊ ”以及“ ㄩˇ ”各有很多同音字，但“国语”两字相连的可能性最大，而当整句输入是“ ㄉㄧㄢˋ ㄋㄠˇ ㄊㄧㄥ ㄍㄨㄛˊ ㄩˇ”时，相对于“电脑听国语”的机率是多少,相对于“店脑听国雨”的机率是多少等，最后会发现“电脑听国语”的机率最高，并把机率最高的句子输出。又例如可以将国小的国语课本的文字，或是报章杂志的文字

（转成电脑档案后）等当作“训练文字”直接输入电脑，电脑就去计算在这些文字中各种不同的字前后相连出现的次数，来建立相当于国小国语课本或某些报章杂志的语言模型。事实上，每一个使用者可以用他自己最适合的训练文字去训练他自己的语言模型；例如财经记者可以用报纸的财经新闻去训练器，则这机器特别适合听写财经新闻，而作家可以用他过去的作品去训练机器，机器则可以适应作家所习用的用语及句型，可用来写稿，错误率可以更低。

上述“中文语言模型”还有一个好处，就是可以部分更正“声音处理器”的错误，因为当两个音十分混淆时，可以一起由“中文语言模型”去选。例如图6 中“ㄍㄨㄛˇ”的机率最高，“ㄍㄨㄛˊ”的机率第二，故应辨认为“ㄍㄨㄛˇ”；但因二者机率接近，可以暂不决定而将两个混淆的音“ㄍㄨㄛˇ”和“ㄍㄨㄛˊ”同样作为候选单音节一起送到后面的语言模型去算前后文的机率，因为下一个音是“ㄩˇ”，“语言模型”会算出来“国语”的机率远比“果雨”高，故最后仍选择了“国语”，错误就被更正了。这种情形和人听国语很像，有些人耳听不清的音，我们会自动根据前后文判断出来是什么音。

这样的“听写机”能听写的字数及词汇端视输入的词典及训练文字的字数及词汇而定。只要输入更多字及词的词典及训练资料，就可将这些数字增大。

以上所说明的是本系列研究最早的第一项申请案中的“中文语言模型”，那事实上是以“字”为基础，亦即计算“字”与“字”相连的机率为最主要的选字参考。但事实上中文文句是以“词”构成，每个“词”是包含了一个到数个“字”，事实上“词”才是中国人造句的基本单位，以图7(a)中的句子为例，该句子可以看成是13个“字”构成的，但是更理想的看法是看成由5 个“词”构成。以此推想，以

“词”为基础的“中文语言模型”，亦即计算“词”与“词”相连的机率为最主要的选字参考，效果一定更好；这也是本系列研究第二项发明在1993年提出申请的基本构想，把上次申请案中以“字”为基础的“中文语言模型”改为以“词”为基础，实验也显示这样的想法是正确的，效果会更好。当时是发展出一种“以词为基础但以字来计算的马可夫中文语言模型”，其说明如图7(b)的例句所示。“今天早上火车站前面人山人海”的例句中共有“今天”“早上”“火车站”“前面”“人山人海”5个词，原应依两两相连计算机率，亦即“今天”接“早上”，“早上”接“火车站”，“火车站”接“前面”，“前面”接“人山人海”等，但当时发展出“以字来计算”的方式，亦即只计算两两相连的词之间相连的字，例如“天”接“早”，“上”接“火”，“站”接“前”，“面”接“人”等。这是因为例如我们可以把所有以“天”结尾的词合成一类，包括“今天”“明天”等；把所有以“早”开头的词合成一类，包括“早晨”“早自习”等，则它们这两类的词两两相连可以都用“天”接“早”来代表，例如“明天早晨”“昨天早自习”等等，故“天”接“早”的机率在此所代表的，事实上是两类更大的词类相连的关系，不仅仅是“今天”和“早上”相连而已。这么一来“词尾字”和“词头字”两两相连的组合仍然只有5千×5千(如果常用字是5千)，故所需的机率值仍然是5千×5千个，和原来以字为基础的语言模型相同；但实验显示它的效果要好很多。此外，当“中文语言模型”是以“词”为基础时，很容易再加入“词频”的信息，也就是越是常用的词越优先选出，这更可进一步提高正确率。

以上所述为本系列研究过去所发展的两种“中文语言模型”的技术，一是以“单字”为基础，计算“字”和“字”两两相连的机率；一是以“词”为基础，但以“词尾字”和“词头字”两两相连的机率

来计算。本发明近年的研究显示，“中文语言模型”的技术发展可以千变万化，可以作出许多种不同的“中文语言模型”，再加以种种组合，可以有非常好的效果。这其中最主要的技术包括：(1) 可以以“字”为单位，以“词”为单位，或以“词群”为单位；(2) 可以计算单独一个“单位”出现的机率，如“字”出现的频率，“词”出现的频率，或“词群”出现的频率；可以计算两个“单位”两两相连的机率，例如两个“字”连在一起，两个“词”连在一起，或两个“词群”连在一起的机率；也可以计算三个单位连在一起的机率；甚至可以计算若干个单位虽不相连但同时出现在同一个句子的机率等；(3) 可以把语言学对中文词类、语法、语意分析所获得的知识或规则和前述(1) (2) 基于统计所获得的语言信息相结合，获得更好的“中文语言模型”。以下将上述三者详述之。

(1) 可以以“字”、“词”、或“词群”为单位。本系列发明最早的方法，也就是图6 的方法，就是以“字”为单位。本系列发明在上项专利中的初步方法，也就是图7(a)中如前所述，计算“今天”接“早上”、“早上”接“火车站”、“火车站”接“前面”等的方法，就是以词为单位。后来在上一项专利中真正使用的方法，也就是图7(b)中如前所述，事实上是把所有以“天”结尾的词合成一群，如“昨天”、“明天”等；所有以“早”开头的词会成一群，如“早晨”、“早自习”等，则这两个“词群”相连都可以用“天”接“早”来代表；故“天”接“早”所代表的还包括了“明天早晨”等其他许多状况。这事实上就是以“某些特性相类似”的词合成一个“词群”，并以“词群”为单位。我们后来的研究显示；把“某些特性相类似”的词合成一个“词群”的方式非常多。除了前述以“同一个字结尾”、“同一个字起头”的词可以合成“词群”以外，文法特性相同的(例如同是及物动词有一个受词的)，语意特性相同的(例如同是指一种

动物的名词），统计特性相同的（例如常常前后接相同的词）等的词都可以合成一个“词群”，而在“中文语言模型”的计算中以这些词群为单位来计算。而前项专利中以“相同字起头”、“相同字结尾”的词构成一个词群的想法，只是这个观念的一个特例而已。图8 显示一个更精细的把中文词分成“词群”的例子。在第一步骤中，以语言学所分析的词类（如动词、形容词、介词）、语意（如“代表状态的动词”如“好像”、“人类角色的名词”如“老师”）、语法（如“有两个受词 的及物动词”如“给”、“由两个名词构成的组合名词”如“台北市长”）知识来把所有的词分成词类一致、语意一致、语法一致的若干群。在第二步骤中，再把第一步骤分好的词类、语意、语法一致的每一个词群中的词，根据在大量文字资料库中的统计特性（例如前接什么样的词、后接那一类的词，或是会和什么样或那一类的词共同出现在同一句中，例如“医生”和“检查”、“警察”和“调查”未必会连在一起，但会出现在同一句中等等），进一步分成统计特性一致的小词群。在第三步骤中，由于第二步骤所分出来词群可能太细了，可以再利用统计特性，把若干个在第一步骤中由于词类、语意或语法不同而分开的小词群但由于事实上统计特性很接近，可以根据统计特性再把它们合并起来。这是一个把词分成精细的“词群”的例子；事实上，很容易想像出来“词群”的分群法千变万化，不同的方法可以得到不同的“词群”，它些“词群”都可以作为中文语言模型的单位，成功的“词群”可以获得成功的“中文语言模型”。事实上“中文语言模型”可以同时包括以“字”为单位、“词”为单位及“词群”为单位的计算，再把计算结果适度组合，获得更好的结果。

（2）可以计算单独一个单位出现的机率，如以“字”、“词”、“词群”为单位时，出现频率高的“字”、“词”、“词群”可以优先被考虑或选出，可以计算两个单位两两相连的机率，如两个“字”

在文字资料库中相连使用的次数，两个“词”相连使用的次数，或两个“词群”相连使用的次数等等，这也是在前两项专利中所使用的。事实上，三个单位连在一起也是很有用的信息，如“我”“要”后面接“去”，“在”“火车站”后面接“前面”等等，故三个单位相连的统计信息也是有帮助的。还有一种信息就是若干个单位虽不相连，但常会同时出现在同一个句子中的，例如“医生”和“检查”、“警察”和“调查”等，这类统计信息也一样可以用在“中文语言模型”中。同样的，各个层次的统计信息，不论是单一单位出现的频率，两个单位两两相连的频率，三个单位相连的频率，乃至虽不相连但同时出现在同一句子的频率等，也一样可以分别计算并适度组合在“中文语言模型”中，使“中文语言模型”的效果更好。

(3) 可以把语言学对中文词类、语法分析所获得的知识、规则或信息和前述基于统计所获得的语言信息相结合，获得更好的“中文语言模型”。事实上，前述图8 中的“词群”分群方法举例，就是一个已经把语言学信息和统计信息结合的“词群”分群方法。其他的例子也很多，例如若前面出现了“把”(这个单音节的同音字不多)，后面八成是“把什么东西作了什么事”，故可以接的词自动减少；若后面出现“了”(这个单音也没有其他同音字)，前面八成是“如何如何了”故前面可以接的词也自动减少等等。

综合前述，“中文语言模型”的技术可以千变万化，把各种技术作各种组合，获致最理想的结果。

至于如何把所获得的“中文语言模型”有效应用在“字词串构成器”所提供的“候选词组”上，搜寻出正确的输出文句，也有多种可以使用的技术。由于单音节辨识未必正确，每一个单音节都可能有好几个候选或混淆单音节，故进入“字词串构成器”时可能有好几个“候选基本音节串”和“候选声调串”；于是由“字词串构成器”所

输出的“候选词组”会包括许多混淆的候选单音节所构成的同音字和同音词，故用“中文语言模型”作输出文句的搜寻时仍不是一件容易的事。这里面可以使用的技巧至少可以包括以下数项：

（1）“声音处理器”所送过来的每一个候选单音节都有它的在“声音处理器”中辨认时的分数，故分数高的单音节所构成的字或词应较优先考虑。

（2）正如前面所说的“中文语言模型”一样，单音节也可视为一种单位，故可计算每一单音节出现的频率，两个单音节两两相连出现的频率，三个单音节相连出现的频率等等；这些频率也可作适度的组合来计算，频率越高的单音节所构成字或词越应优先考虑等等。

这些都是在应用“中文语言模型”在候选词组中搜选输出文句的重要技术。

综合上述各部分，就是“语言解码器”的核心技术。

以上是说明了“声音处理器”和“语言解码器”，这两项是本发明两项最基本的技术。这两项基本技术使得本发明的能力向前迈进一大步，可以处理的输入语音将不再限于“断开的单字”，也可以是“断开的词，词中的字是相连的”、“断开的音韵段，段中的字是相连的”，甚或是“整句完全连续的”国语语音。图9 以一句“今天早上我在火车站前面遇到我的老师”为例，说明这四种输入方式在使用者念的时候的区别。以下再说明本发明进一步发展出来的许多“智慧型学习技术”，也画在图1 的下半部中，使得本发明的听写机具备不时“学习”的“智慧”。

第一项学习技术是以分阶段的“学习例句”来阶段性的自动学习使用者的声音，亦即是用一系列几个阶段特别设计的“学习例句”。新的使用者只要念最前面第一阶段的若干句，即可使机器初步学习会听使用者的声音。这是因为这若干句共包含了所有的所使用的国语的

“次音节单位”。例如念了一个“ ㄅㄚ ”音，机器也会同时学到“ ㄅㄠ ”“ ㄅㄢ ”等音的声母，“ ㄅㄨ ”“ ㄇㄚ ”等音的韵母等；故这些句子念下来已可学到所有可能的声音。同时这些句子中也让越常出现的“次音节单位”出现次数越多，所以训练得越正确。这是为什么只要很少句就可以初步学会使用。在第二阶段中，若新使用者愿再多念若干句，就可以把正确率再大幅提高。这是因为这些第二阶段的句子中已把常用的200 个国语基本单音节全部纳入，并让越常用的单音节出现次数越多，训练得越好。例如“ ㄊㄚ ”是一个常用的音，在最前面的第一阶段的“学习例句”中，这个音是靠“ ㄊㄠ ”的声母和“ ㄆㄚ ”的韵母训练的，所以不是很精确，但在这第二阶段的若干句中，“ ㄊㄚ ”会多出现几次，所以会训练得不错，正确率也大幅提高。在第三阶段中若新使用者愿意再多念若干句，则正确率可以再提高，其原因与上述相同，只是此时第三阶段的“学习例句”包含了所有国语中可以出现的400 多个基本单音节(但不计声调) ，且常出现的也多念几次。在第四阶段中，若新使用者愿意再多念若干句，即可把正确率再为提高，这是因为在这些句中，所有国语的1300个单音节包括不同的声调(例如“ ㄅㄚ ”事实上有5种变化：“ ㄅㄚ ”“ ㄅㄚˊ ”“ ㄅㄚˇ ”“ ㄅㄚˋ ”“ ㄅㄚ· ”) 都会念到一次以上，且越常用的出现次数越多。

第二项学习技术是机器自动“线上”学习使用者的声音。使用者事实上不必做完上述的四个阶段的学习才开始使用机器，而是可以在作完上述第一阶段的训练以后的任何时候开始使用，只是正确率较低而已。不论是用上述的各阶段“学习例句”训练机器，或是在真正使用中，只要随时更正错误，机器立刻作“线上学习”，亦一面使用中一面把所有辨别过的声音全部学习进去，因此只要使用者继续使用并让机器学习，正确率可以逐步达到95%－97%左右，亦即约每20-35

字才须修正一个错字。

第三项学习技术是线上自动学习环境噪声。每一个使用者的环境都有他自己的噪声，这些噪声都会对机器的使用正确率造成伤害。在本发明的上述第二项“线上学习使用者的声音”的过程中，事实上机器还可以自动学习使用者的环境噪声的特性。并适应之。因此学习一段时间以后，机器就可以在环境噪声下工作得很好。

上述三项学习功能所使用的技术事实上是相同的。首先先用很多位不同的说话者 所发的声音，来训练国语每一个“次音节单位”（不论是选择那一种“次音节单位”）以及“声调模型”的“隐藏式马可夫模型”。因为很多位不同的说话者声一定不同，即使是发同一个“次音节单位”，也会有相当大的不同，故这样多说话者的“次音节单位模型”及“声调模型”的“隐藏式马可夫模型”中，常常需要相当多数目的高斯机率混合，才可以涵盖不同的说话者发这一个单音的各种不同的声音特性。当新使用者念这一个“次音节单位”及“声调”的时候，就用一套演算法去在多说话者的“隐藏式马可夫模型”的许多高斯机率混合中找出最接近新使用者声音的那几个高斯机率混合，而把其他的高斯机率混合抛弃，这时的“隐藏式马可夫模型”就会变成新使用者的“隐藏式马可夫模型”了。以后新使用者的声音继续进来，可以再把新的声音加进去一起平均算出新的高斯机率混合，于是新使用者声音的成分越来越多，这个“隐藏式马可夫模型”（包括“次音节单位”或“声调模型”）就越来越能精确的描述新使用者的声音，正确率也就越来越高。当使用者的环境有噪声时，噪声夹着新使用者的声音一起进来，也会一起把噪声的特性平均进去，因此所算出的高斯机率混合就自动带着噪声特性作为背景了。因此所训练出来的“隐藏式马可夫模型”就自动能适应该种特性的噪声了。值得一提的是我们也成功的发展出“隐藏式马可夫模型”的十分简化的数学

架构，演算十分方便快速，因此才可以作“线上”学习；也就是使用者一面使用，一面声音就被平均进去，下一次念的时候就是用新的模型来辨认，因此“线上学习”的效果可以很快而显著。

第四种学习技术是线上自动学习使用者的用字，用词及构句习惯，每一个使用者基本上都会有他自己特别的用字、用词及构句习惯，事实上很多错误发生是因为机器不能学习使用者的这些习惯。因此当使用者一面使用机器，并将错误作线上更正后，机器立刻把使用者用过的文句，包括里面的用字、用词及构句学习进去，也就是把诸如词频，两两相连的机率等语言模型的重要参数重新计算一次并调整之，于是机器就学到了使用者的用字，用词及构句习惯。

第五种学习技术是短期储存保留。在输入一段文字时，当这段文字在讨论某一事物，若干特别的用词，构句常会重覆出现，此时经线上更正后，机器可以把这些特别的信息包括词频，两两相连的机率等保留在短期存储器中优先参考使用，因此越用到后来正确率会越高。当改输入其他主题的文字时，这些短期存储器中的信息可以全部消除。

以上第四、五两种学习技术详细情形之一举例请见图10。当“声音处理器”送过来一串辨认出来的可能的音节串时，先藉助词典及“字词串构成器”查出所有可能的字及词串，再用“中文语言模型”及“词频”等信息找出最可能的句子输出。使用者可以作线上更正，机器就会立刻学习。学习包括“长期学习”和“临时学习”两种。“长期学习”也就是算出新的词频及“中文语言模型”中的新的机率等，而“临时学习”则包括可以建立一个临时新词典存放一些新词并包括这些新词的词频。这个新词典及新词频在输入这篇文章结束以后，使用者可以决定并入整个词典及词频信息中，也可以将之取消。此外，也常有一些用词或构句是这一篇文章在讨论某一事物时特别会重覆出现。若仅学习进入整体词典及整体“中文语言模型”中，学习效果并

不明显，因为这些用词或构句也不过多出现几次，对整体的词频及两两相连的机率等影响不大。因此在本发明也可以另外建立一个短期储存，如图10下方，里面存有为这篇文章所特别计算的词频及两两相连的机率等；机器在寻找句子时，优先在短期存储器中找寻答案，找不到时才诉诸整体模型及整体词典词频。这样这篇文章特有的用词，构句就会被学会，因此越输入到后面，正确率会越高。但等到下次输入主题不同的另一篇文章时，此一短期储存可以全部清洗掉，故不致干扰后面的输入工作。

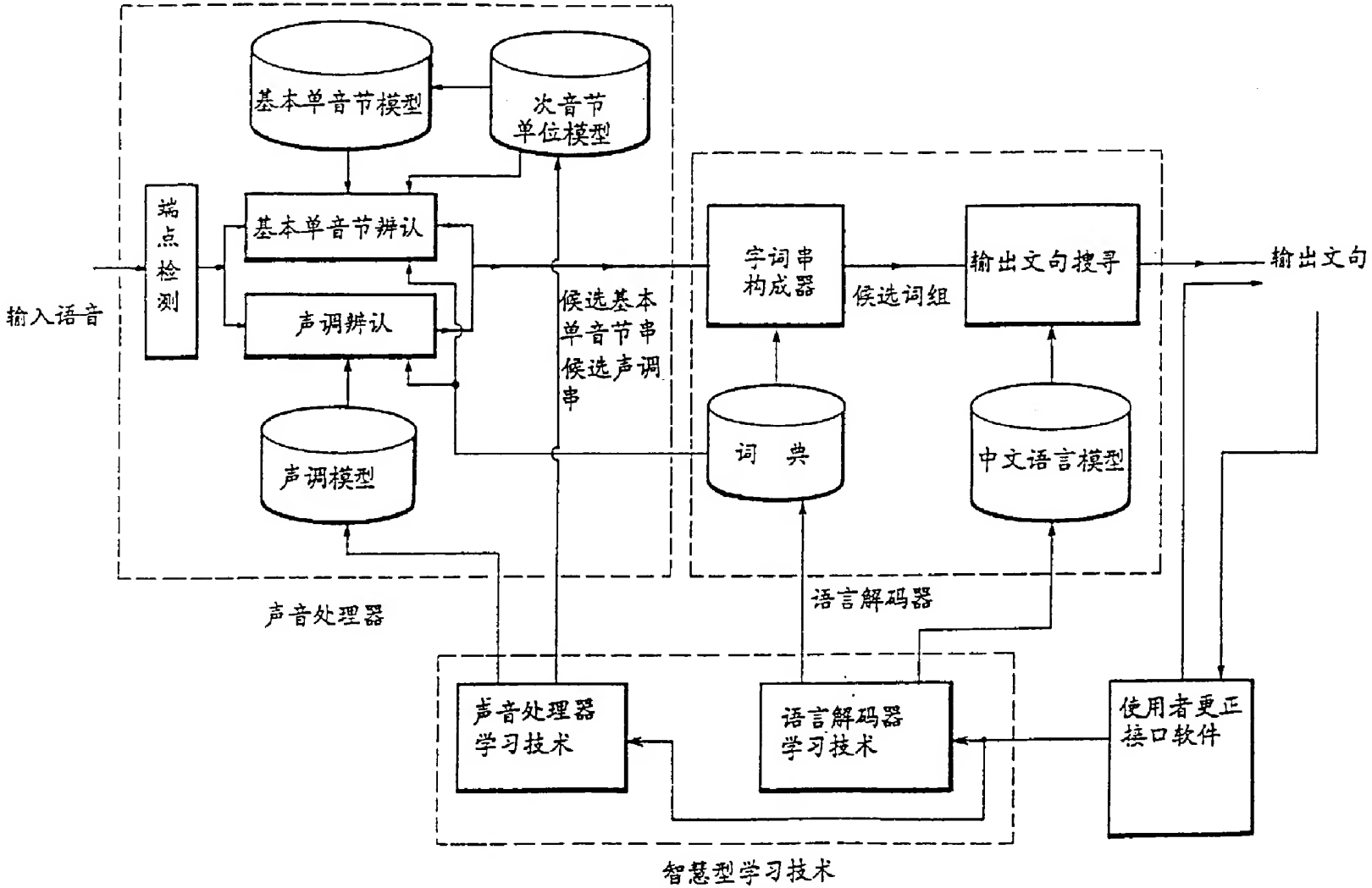
本发明中尚有几项技术需被充说明。第一项是用计算机程式来自动选取“学习例句”的技术。如前所述，本发明有一套特别设计的“学习例句”，新的使用者因此只需念最少的句子就可以训练机器听他的声音。这些特别的“学习例句”事实上是由电脑在一大堆文章档案中搜寻出来的，图11是这样一个电脑自动选句的演算法的流程图举例。其基本原理是把所有想要的基本单位音（声母、韵母、次音节单位、声韵母相连、声调、单音节、基本单音节等），都可以给定分数；而文章档案中的每一句子也可根据句中所包含的基本单位的分数算出句子的分数；当然同一句中若含越多不同的基本单位音，就分数越高，因此就越优先被挑出来；可是一个句子一旦被挑出，它所有包含的基本单位音的分数就自动归零，也就是下次不再优先选出包含这些已出现过的基本单位音的句子了。此外，为了让平常出现越多（也就是越常用）的基本单位音在“训练例句”中也出现越多次，以使训练得更精确，因此利用一个参数来描述各个基本单位音出现的频率分布和它们在正常用语中真正的频率分布接近的程度，故可用这个参数来选句，以致于只用很少的句子就可以使得越常用的音出现越多，也就是频率分布越接近真实情形。

另一项技术是“动态词典结构”。由于词典中词的数目极为庞大，

每次搜寻耗费时间甚多；其中尤其单字词，双字词特别多。因此本发明设计出“动态词典结构”，也就是把最常用的双字词，单字词找出来，加上其他的三字以上的长词，构成一个“常用词典”，其他的词则放在另一个“罕用词典”中。机器操作时原则上只在“常用词典”中找词，找不到词无法构成理想句子时才去“罕用词典”找。在“罕用词典”中找出来而正确的词学习后就放入“常用词典”中，而“常用词典”中的词若久不使用，也可移入“罕用词典”。如此在词典中找词所费的时间，可以缩减到约1/10。

图12为本发明的一个较佳具体实施例。主机是一台个人电脑，而整个国语听写机的技术则可完全以软件完成，写入一片Ariel DSP 96003D数位信号处理电路板上，包括所有“基本单音节模型”、“次音节单位模型”、“声调模型”、“中文语言模型”，许多的演算技术，以及前述的许多智慧型学习技术作业，全部写在这片电路板上。所有的运算只靠电路板上的一片数位信号处理晶片Motoralla DSP-9600即可完成。事实上市面上可以选用的数位信号处理晶片及电路板很多，本较佳具体实施例所用的只是本发明在台大实际制作时所用的例子而已。使用者的声音由麦克风输入电路板，听写机完成听写程序后，把中文字显示在个人电脑的荧光幕上。

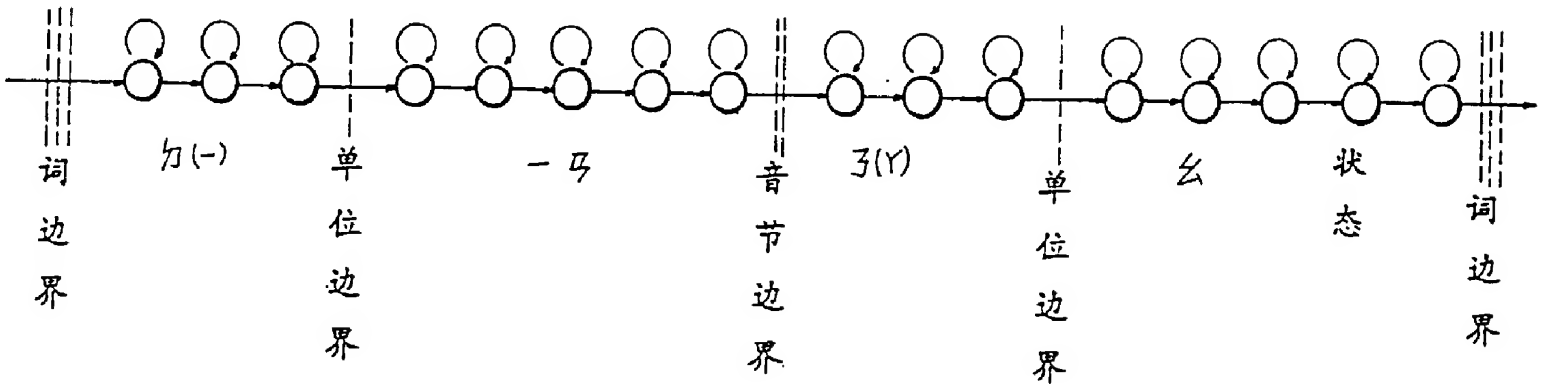
前述的实施例只是用以说明本发明的原理，并不能用此限制本发明。任何人依据本发明的原理所做的修改皆应仍隶属于本发明的精神。本发明的范畴当如后列的权利要求书所列。



本发明的基本原理及技术架构图

图 1

(a) “受后接韵母起始音素影响的声母”和“不受前后音影响的韵母”



(b) “受后接音素影响的音素”

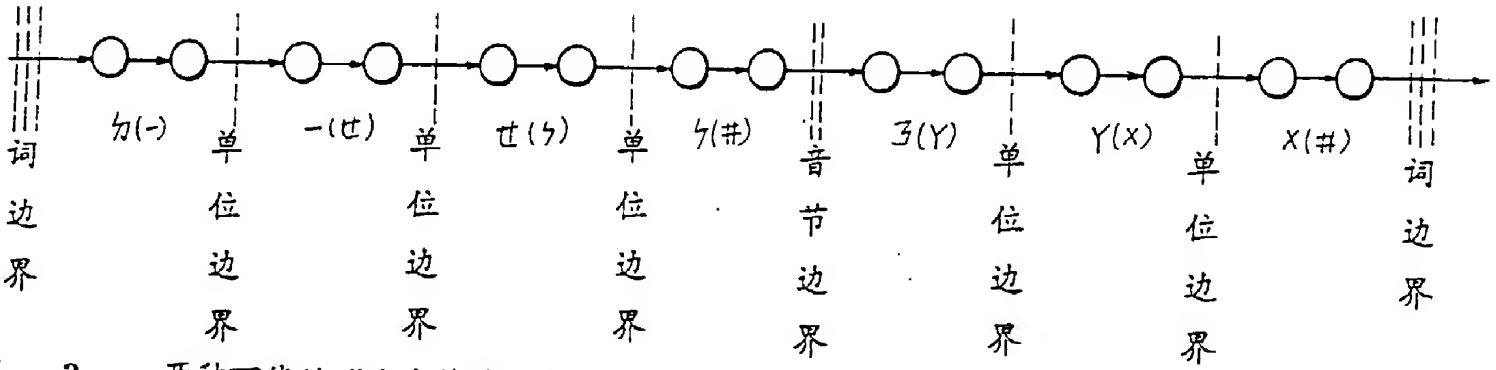
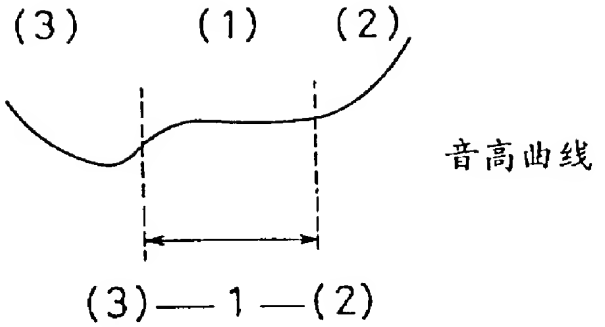


图 2 两种可能的“次音节单位”举例，均以“电脑”一词的基本单节（ㄅ一ㄢ，ㄋ幺）为例说明

(a)



(b)

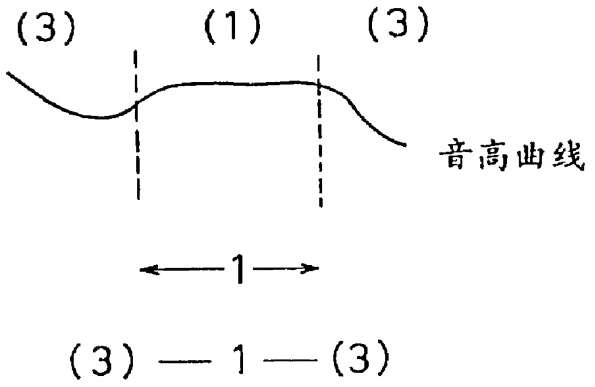
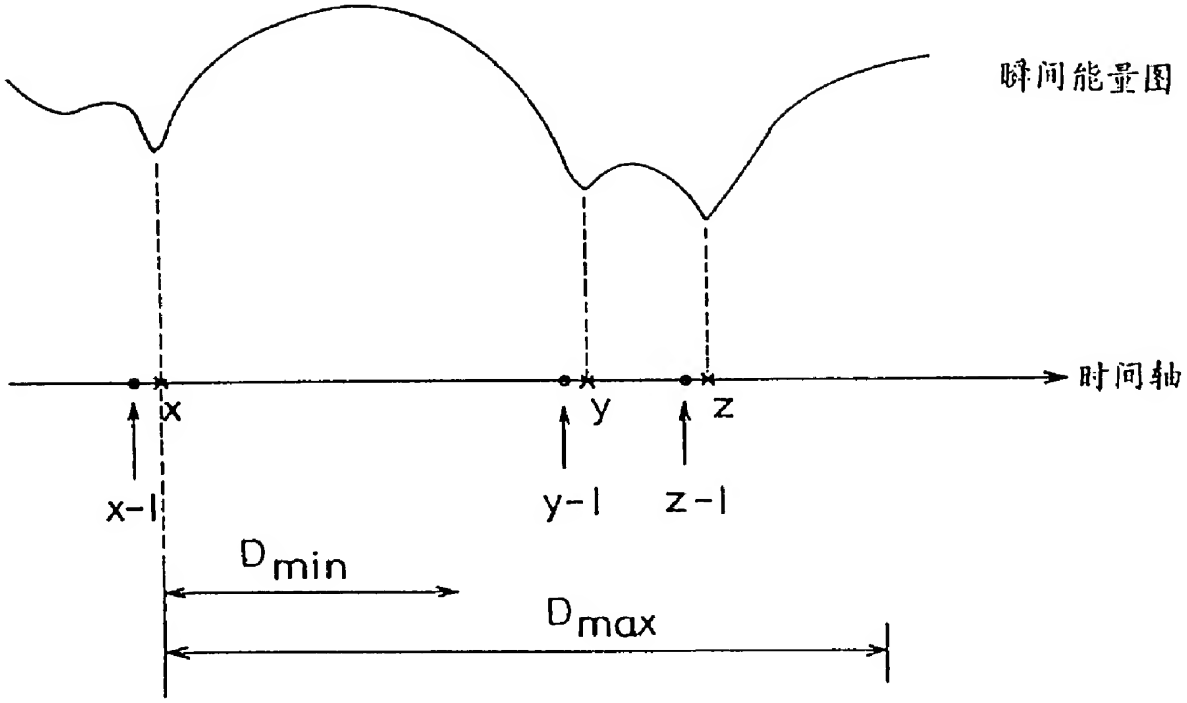


图 3 “受前后声调影响的声调模型”

(3)-1-(2)表示“前接第三声，后接第二声”的第一声等等，在图中因后接第二声或第三声对音高曲线的影响差别不大，(3)-1-(2)和(3)-1-(3)可以合并成为一个模型等等。



x, y, z ： 可能的音节起点

x-1, y-1, z-1 ：可能的音节终点

D_{min} , D_{max} ：分别为一个音节可能长度的下限及上限

$$T[y-1] = T[x-1] + \max_{i} S(x, y-1)$$

图 4 连续音节比对法

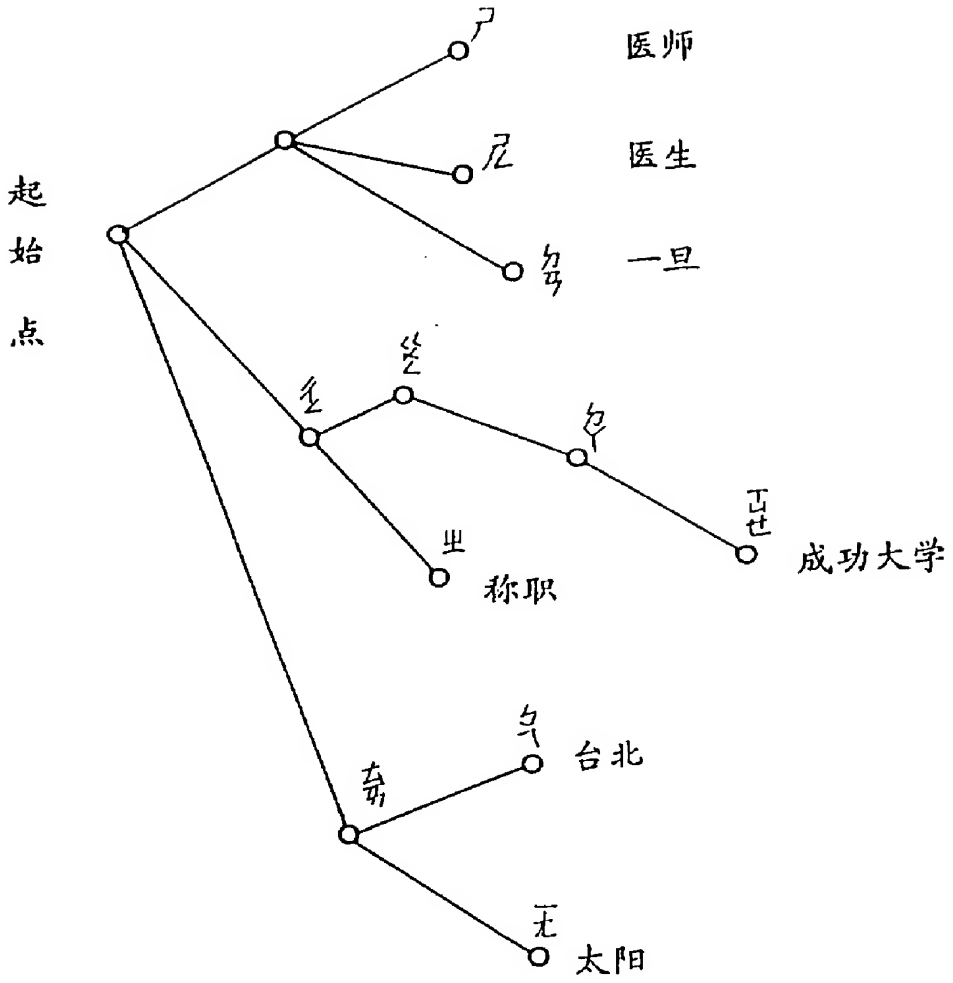
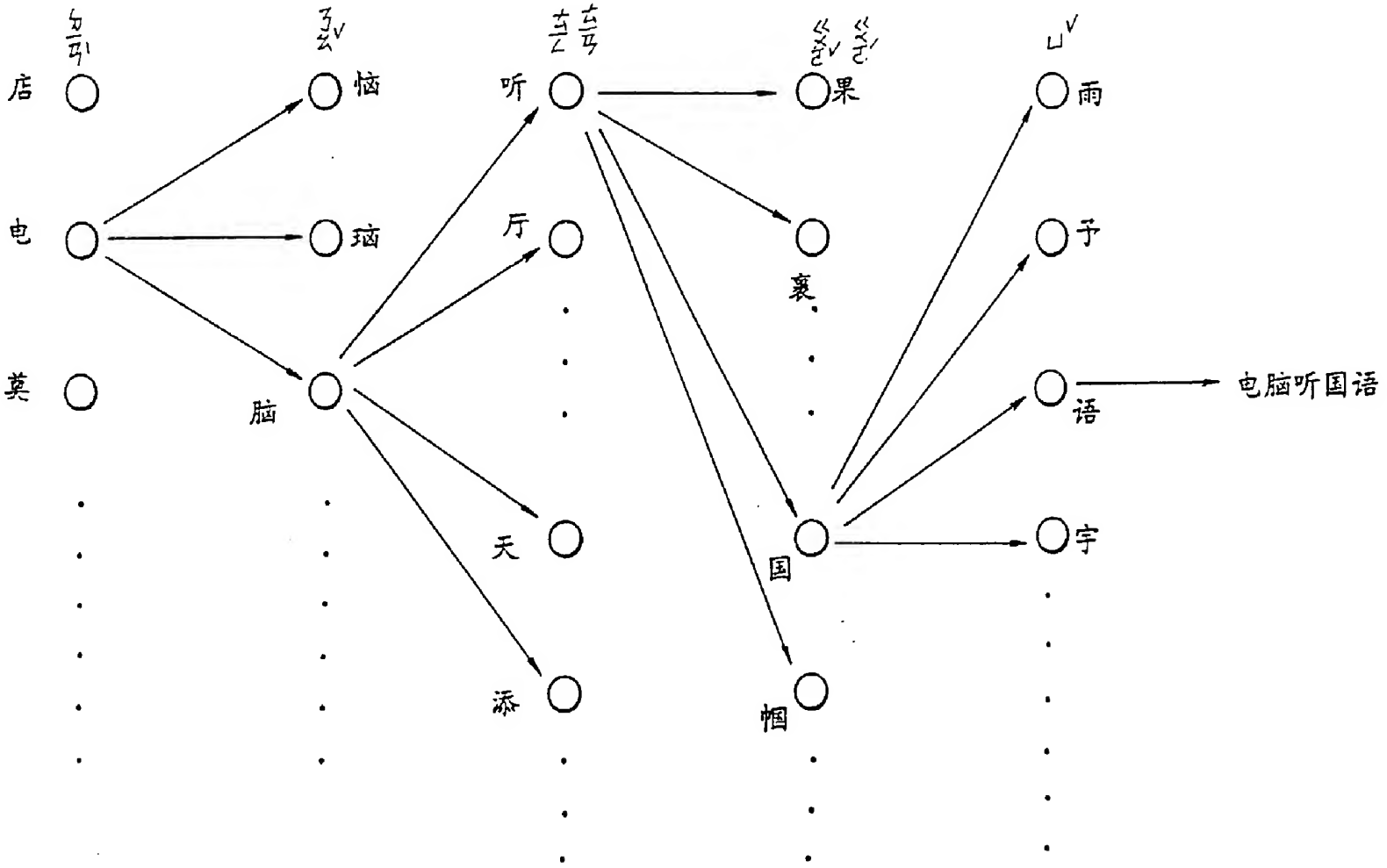
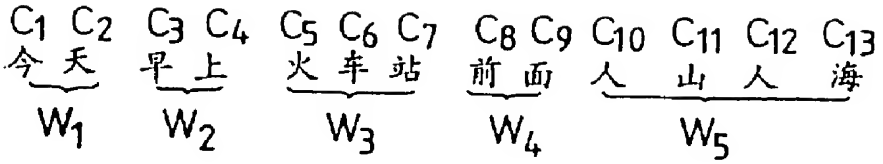


图　5　树状词典资料结构

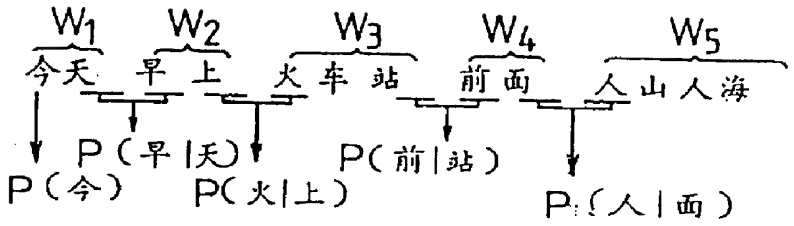
图 6 以字为基础的“马可夫中文语言模型”



(a)



(b)



$$P(W) = P(W_1, W_2, \cdots, W_m)$$
$$= P(\underbrace{C_{11}C_{12}\cdots C_{1S_1}}_{W_1}, \cdots, \underbrace{C_{i1}\cdots C_{iS_i}}_{W_i}, \cdots, \underbrace{C_{m1}C_{m2}\cdots C_{mS_m}}_{W_m})$$
$$= P(C_{11})P(C_{21}|C_{1S_1})\cdots P(C_{m1}|C_{(m-1)S_{m-1}})$$
$$= P(C_{11}) \cdot \prod_{i=2}^{m} P(C_{i1}|C_{(i-1)S_{i-1}})$$

图 7 以词为基础但以字来计算的马可夫中文语言模型

(a) 以语言学分析的词类、语意、语法知识分群

(b) 以文字资料中词与词前后相连或同时出现的统计特性进一步分群

(c) 有些统计特性接近的群，但在第一阶段(a)中因为词类、语意、语法不同而被分开者，可以再予合并

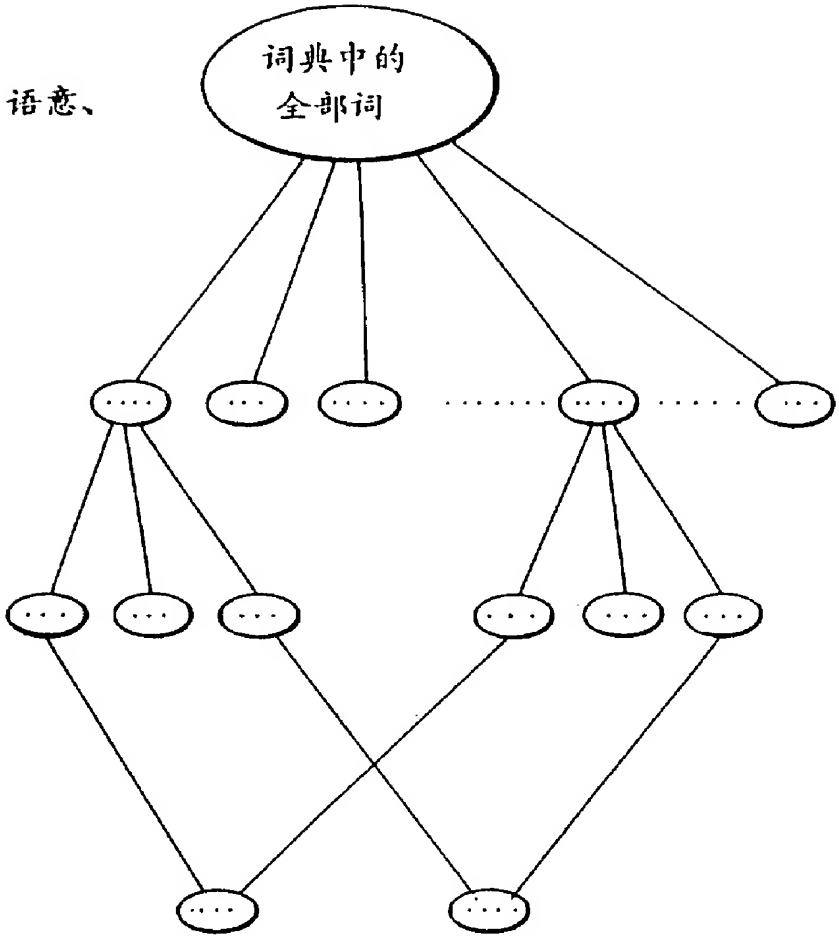


图 8 结合统计特性及词类、语意、语法等语言学知识或规则来作词群的方法举例

(a) 今，天，早，上，我，在，火，车，站，前，面，
遇，到，我，的，老，师，

(b) 今天，早上，我，在，火车站，前面，遇到，
我，的，老师

(c) 今天早上，我在火车站前面，遇到我的老师

(d) 今天早上我在火车站前面遇到我的老师

图 9 各种可能的国语语音输入方式:

(a) 单字为单位 (b) 词为单位 (c) "音韵段" 为单位
(d) 整句连续输入

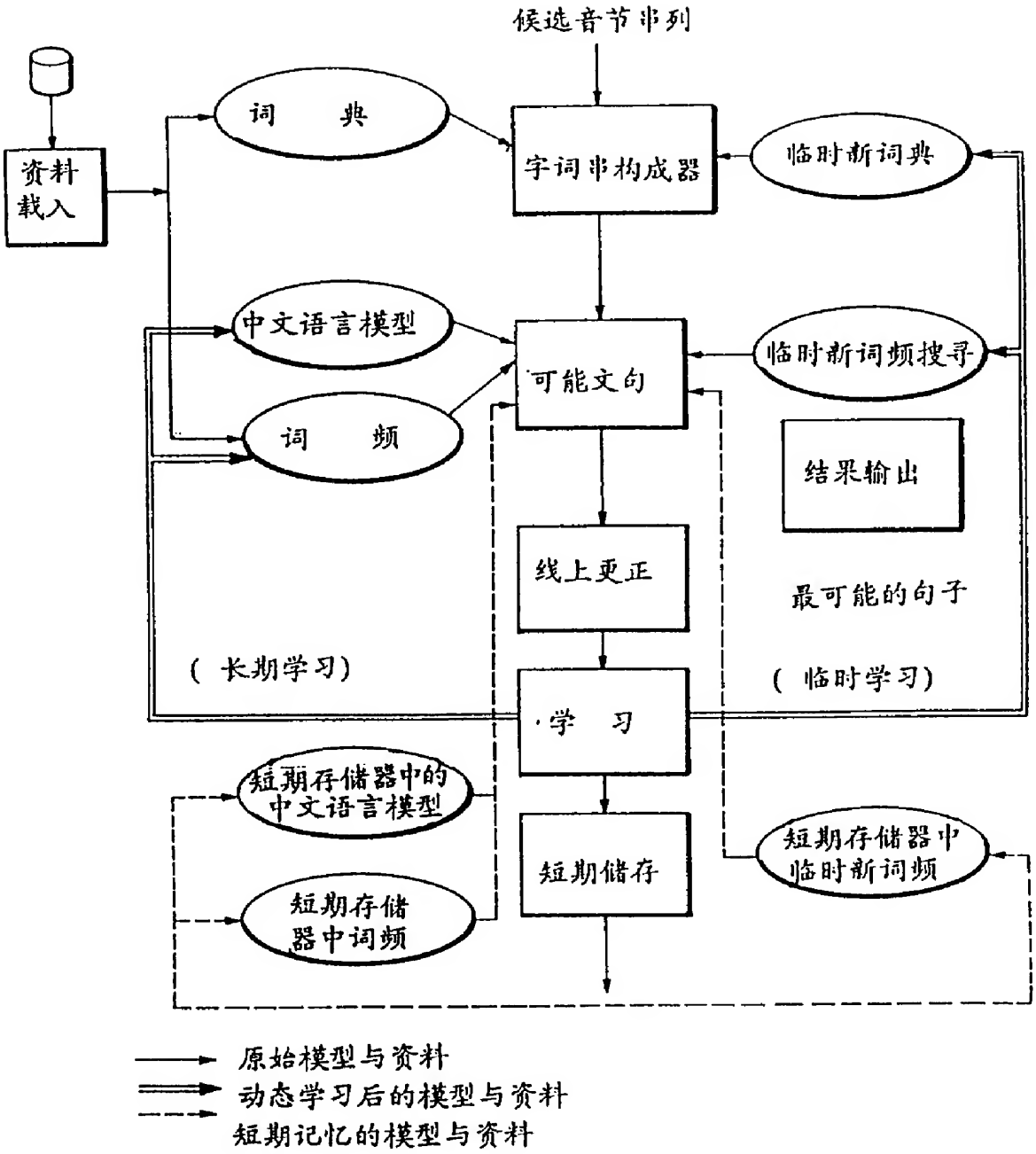


图　1.0 “语言解码器”的智慧型学习技术可能作法的细节举例

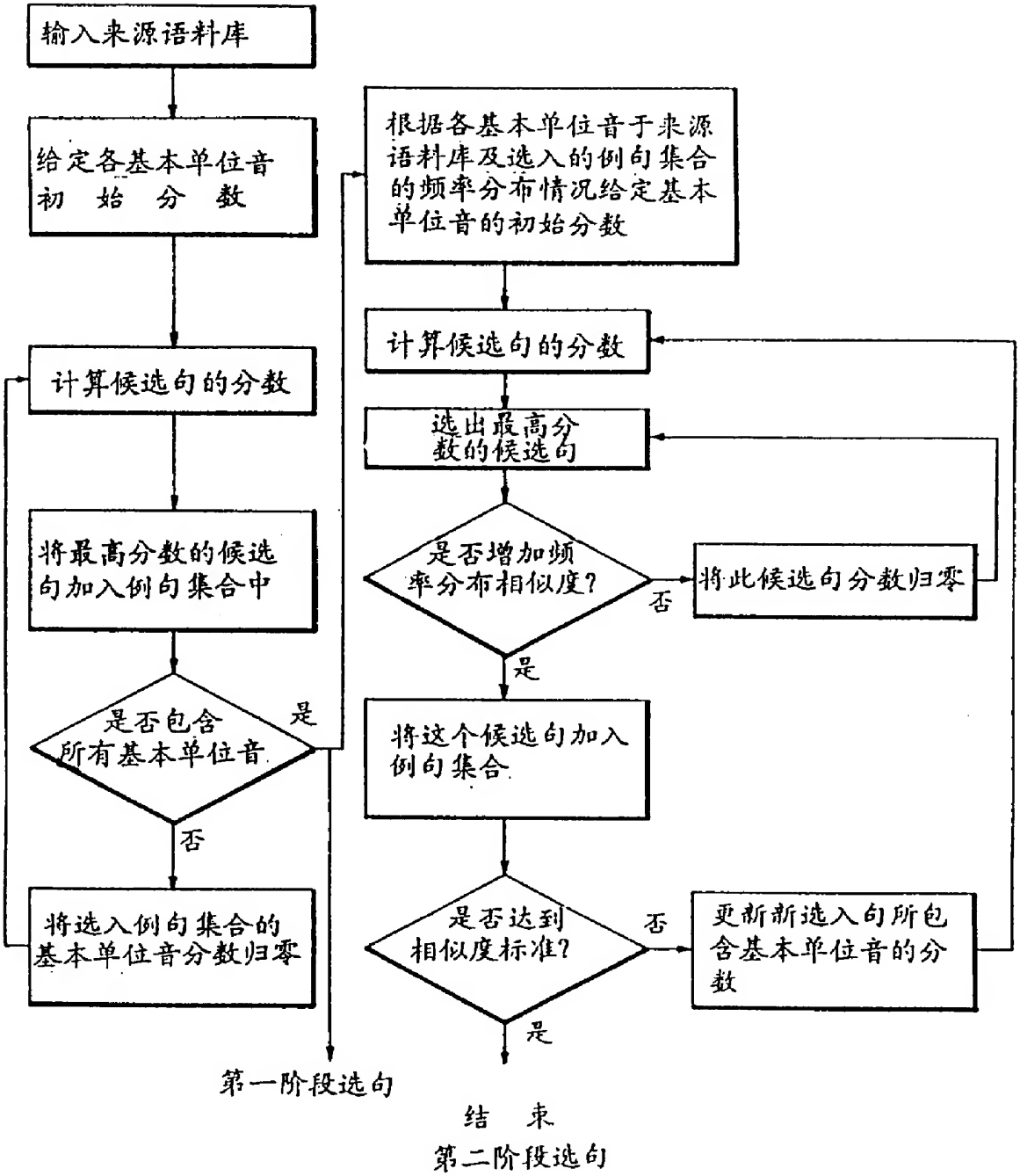


图 11 电脑自动选取“学习例句”的方法

具 体 实 施 例

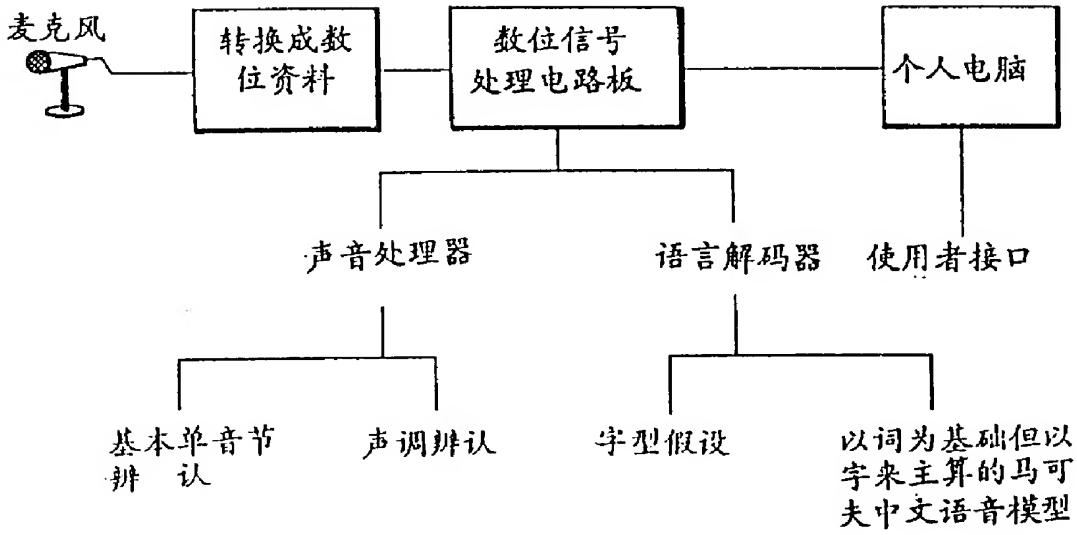


图 12 本发明的一个较佳具体实施例